DOCUMENT RESUME

ED 329 313 JC 910 147

AUTHOR Scott-Skillman, Thelma; Halliday, Karen
TITLE Matriculation Assessment Standards.
INSTITUTION California Community Colleges, Sacramento. Office of the Chancellor.
PUB DATE Mar 91
NOTE 26p.; Discussed as agenda item 6 at a meeting of the Board of Governors of the California Community Colleges (Stockton, CA, March 14-15, 1991).
PUB TYPE Legal/Legislative/Regulatory Materials (090)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Standards; Community Colleges; Educational Legislation; *Minimum Competency Testing; *Standardized Tests; State Standards; Statewide Planning; *Student Placement; Testing; Test Norms; Test Selection; *Test Validity; Two Year Colleges
IDENTIFIERS *California Community Colleges

ABSTRACT

Assessment testing in California's community colleges (CCC) helps to identify entering students' goals and skills level and to ensure that the students are properly advised of appropriate courses and programs. The California Education Code requires that all assessment instruments be approved by the Office of the Chancellor of the California Community Colleges. As part of its efforts to establish methodologies for use by colleges in reviewing their placement instruments, the Chancellor's Matriculation Assessment Work Group developed a document to guide the choice and use of assessment instruments. Entitled "Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges," this four-part guide: (1) provides a definition of matriculation, reviews regulations on assessment procedures as contained in Assembly Bill 3, presents an outline of the seven-step process for recommending specific assessment instruments to the Chancellor, lists provisions and prohibitions pertaining to specific matriculation services used in the CCC, and summarizes the rationale for the use of assessment for placement purposes; (2) reviews standards for the evaluation of assessment measures, examining issues of test validity; reliability and errors of measurement; scaling, norming, score comparability and equating; standards for administration, scoring and interpretation; and testing special groups; (3) discusses test usage criteria specific to the California community colleges; and (4) details the seven-step process for reviewing assessment measures. (GFW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED329313

MATRICULATION ASSESSMENT STANDARDS

6

First Reading, Action Scheduled

Staff Presentation: Thelma Scott-Skillman
Vice Chancellor, Student Services

Karen Halliday
Coordinator, Matriculation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

JC 910 147

Board of Governors
California Community Colleges
March 14-15, 1991
Stockton, California

BEST COPY AVAILABLE

Board of Governors
California Community Colleges
March 14-15, 1991

MATRICULATION ASSESSMENT STANDARDS 6

First Reading, Action Scheduled

Background

Matriculation is a process that enhances access to the California Community Colleges and promotes and sustains the efforts of credit students to succeed in their educational endeavors. The goals of matriculation are to ensure that all students complete their college courses, persist to the next academic term, and achieve their education objectives.

Assessment is a major component of the matriculation process. It helps identify the entering student's goals and skill levels and ensure that he or she is properly advised of appropriate courses and programs. While assessment employs a variety of methodologies and measurements, assessment testing is one method common to all of California's community colleges.

Section 78210 of the *Education Code* provides that:

> (a) No district or college may use any assessment instrument for the purposes of this article without the authorization of the Chancellor. The Chancellor may adopt a list of authorized assessment instruments pursuant to the policies and procedures developed pursuant to this section and the intent of this article. The Chancellor may waive this requirement as to any assessment instrument pending evaluation.
>
> (b) The Chancellor shall review all assessment instruments to ensure that they meet all of the following requirements:
>
> (1) Assessment instruments shall be sensitive to cultural and language differences between students.
>
> (2) Assessment instruments shall be used as an advisory tool to assist students in the selection of an educational program.
>
> (3) Assessment instruments shall not be used to exclude students from admission to community colleges.
>
> (c) The Chancellor shall establish an advisory committee to review and make recommendations concerning all assessment instruments used by districts and colleges pursuant to this article.

(Note: The text of Section 78210 is displayed as it read up until January 1, 1991. As a part of the Board's Education Code Review bill, references to the "Chancellor" have been changed to the "Board of Governors.")

A Matriculation Assessment Work Group was established by the Chancellor to develop standards and procedures for evaluating the assessment (placement) instruments used to advise students in course selection. The work group includes members from the Chancellor's Office, the districts and colleges, and the Center for Educational Testing and Evaluation at the University of Kansas (contracted specifically for this project).

During 1990, the work group developed *Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges* (Appendix). In addition, the work group has established methodologies that the colleges themselves will use when conducting their own reviews of placement instruments. In conjunction with college staff, this work group has begun the process of identifying, assembling, reviewing, and evaluating the instruments currently used in the community college system. The Chancellor plans to establish an initial list of approved instruments by July 1, 1991.

A *Matriculation Assessment Technical Manual* to assist districts and colleges in understanding the psychometric issues related to test standards was distributed to the field in August, 1990. Also, eleven research studies to assist colleges in validating instruments were done by contracted community college researchers. These were distributed in January, 1991.

Analysis

Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges contains four major sections. The first catalogues and details the purposes and appropriate uses of assessment as defined by Title 5 regulations. The second presents and discusses the standards and criteria to be used in the review of assessment instruments. The psychometric standards are: (1) validity; (2) reliability and errors of measurement; (3) scaling, norming, score comparability, and equating; (4) standards for administration, scoring, and interpretation; and (5) testing special groups.

The third section of *Standards, Policies and Procedures* presents more explicit standards and criteria for evaluating assessment instruments. These include content and predictive validity; reliability and errors of measurement; analysis for cultural/linguistic bias, insensitivity, and offensiveness; cut-off scores, and the impact of testing on various groups.

Specific criteria are being developed for four types of assessment instruments: (1) multiple-choice and other objective achievement tests developed by an external agency (not associated with the California Community Colleges); (2) multiple-choice and other objective achievement tests developed by community colleges; (3) writing samples and other measures with open-ended responses; and (4) non-achievement instruments used for placement, such as interest measures and inventories. These standards help district and college staff assess the quality of all types of assessment instruments and to describe the obligations of all participants involved in the testing process, including test developers and users.

The fourth section of *Standards, Policies and Procedures* details the assessment-instrument review process that leads the Chancellor's recommendation regarding its uses. This is a seven-step process:

Step 1. Compile information on assessment instruments.

Step 2. Perform psychometric expert review.

Step 3. Perform content expert review.

Step 4. Perform review by Matriculation Assessment Work Group. This step also involves discussion of the preliminary findings with the test publisher.

Step 5. Generate recommendations (approval, provisional approval, probationary approval, or disapproval).

Step 6. Disseminate Chancellor's decision.

Step 7. Identify appeals process.

In conclusion, the standards, procedures and policies proposed by the Matriculation Assessment Work Group are consistent with legislation and regulations. They enable the Chancellor to appropriately evaluate the assessment instruments used by districts and colleges for their validity, reliability, and sensitivity to language and cultural differences.

Recommended Action

That the Board of Governors adopt the document, *Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges*, developed by the Matriculation Assessment Work Group.

Staff Presentation: *Thelma Scott-Skillman, Vice Chancellor*
Student Services

Karen Halliday, Coordinator
Matriculation

APPENDIX

Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges

PREFACE

Matriculation is a community college process that promotes and sustains the students' efforts to achieve their educational goals. A focus for matriculation is on affording the individual access to educational opportunities, and then taking steps to increase the likelihood of the individual's success. One service of matriculation is to provide assessments that are useful for assisting the student 's selection of an educational program. That assessment instruments are fallible is axiomatic, and thus the goal is to select and then use instruments that provide the most accurate and useful information. This document describes the standards, policies and procedures that should guide the choice and use of assessment instruments employed in the California community colleges.

SECTION I: AN INTRODUCTION TO ASSESSMENT IN MATRICULATION

Regulations to implement the California legislative mandate known as "matriculation" (AB3) define matriculation as

> **a process that brings a college and a student who enrolls for credit into an agreement for the purpose of realizing the student's educational objectives through the college's established programs, policies, and requirements.**

The intent of AB3 is to establish a matriculation system that includes minimum standards in the California community colleges and provides guidelines for implementing these standards to make certain that adults pursuing postsecondary education have equal access to programs and services and opportunities for success.

One major component of the matriculation process is assessment. AB3 and its regulations clearly indicate that the primary function of such assessment is to assist the student in making decisions about appropriate course level enrollment, major area of study and vocational program choice. Assessment's primary role in matriculation is viewed as providing descriptive and predictive information about students and their "fit" to courses and programs, thus facilitating their potential for success at the community college.

The intent of the legislation, where assessment is concerned, is to establish guidelines, procedures and standards for ensuring that assessment instruments and procedures implemented in the California community college system are appropriate and in line with intended use as defined and restricted by AB3 and its matriculation regulations. As such, the act specifically requires that an advisory committee to the Chancellor of the

BEST COPY AVAILABLE

California community colleges review and make recommendations concerning all matriculation assessment instruments used by the colleges. Based on these recommendations, the Chancellor shall establish and update, at least annually, a list of approved assessment instruments and guidelines for their use by community college districts. In line with this requirement, the purpose of this document is to: (1) specify the procedures for the advisory committee to follow in arriving at these recommendations, (2) delineate the standards that the committee should consider in the review and evaluation of assessment instruments, and (3) define the information that shall be provided to the Chancellor by the advisory committee concerning the assessment instruments.

The recommendations that are formed are to be established based on those professional standards that guide educational and psychological testing and those conditions called for in AB3 and its regulations. While the review will focus on specific instruments (tests), one should understand that the final scrutiny will be on the suitability and appropriateness of the use, i.e. the tests' scores' interpretation and resulting recommendation(s) must be evaluated and ultimately judged. The final responsibility for the proper use of assessment instruments and procedures and resulting scores remains with a local college. An affirmative recommendation by the Chancellor regarding a test only provides the opportunity for a district to consider its use. An affirmative recommendation does not automatically endorse the local college's use of test scores as proper.

Keeping this stipulation in mind, a seven-step process will be followed for making recommendations to the Chancellor about specific assessment instruments. These seven steps are detailed in a later section, but include the following:

Step 1. Compile Information on Assessment Instruments
Step 2. Develop Psychometric Expert Review
Step 3. Develop Content Expert Review
Step 4. Develop MAC Assessment Work Group Review
Step 5. Generate Psychometric Expert Recommendations
Step 6. Disseminate Chancellor's Decision
Step 7. Identify Appeals Process

For assessment instruments approved by the Chancellor, each community college district must still develop documentation to demonstrate that it is using the test appropriately. This information will be periodically evaluated during site visits by a team representing the Chancellor's Office.

If the Chancellor disapproves an assessment instrument, no community college district may use the test except on an experimental or pilot basis and then only with the permission of the Chancellor's Office. The purpose for experimental use is to collect information pertinent to a re-review of the

instrument in an attempt to obtain future approval for its use from the Chancellor's Office. Local users, in these instances, would place emphasis in their documentation efforts on areas defined as deficient in the report filed at the time a negative decision was made concerning a test's use. A re-review of an instrument would involve all seven steps in the process.

The remainder of this document delineates the procedural steps defining the assessment instrument review process and identifies the reviewers' criteria (standards) used when examining an instrument. These standards define the criteria by which judgments will be made about the recommended acceptability of an instrument's use by the psychometric expert review committee to the Chancellor as well as the evaluation of instruments in use by colleges during site visits by a team representing the Chancellor's Office.

Assessment in Matriculation

Where assessment by instrument is concerned, the matriculation Title 5 regulations place responsibility for the approval of assessment instruments with the Chancellor of the California community colleges. This responsibility necessitated the development of review procedures and criteria for making the determination. Specifically, the regulations state that:

> **The Chancellor shall establish and update, at least annually, a list of approved assessment instruments and guidelines for their use by community college districts. These guidelines shall identify modifications of an assessment instrument or the procedures for its use which may be made in order to provide special accommodations required by Section 55522 without separate approval by the Chancellor. Such guidelines shall also describe the procedure by which districts may seek to have assessment instruments approved and added to the list. The Chancellor shall ensure that all assessment instruments included on the list minimize or eliminate cultural or linguistic bias, are normed on the appropriate populations, yield valid and reliable information, identify the learning needs of students, make efficient use of student and staff time, and are otherwise consistent with the educational and psychological testing standards of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. (Section 55524)**

Within the matriculation regulations, the broader implications of this latter directive are presented through the definitional and clarifying statements on what "assessment instruments, methods and procedures" encompass. In the regulations, assessment is defined as:

> **the process of gathering information about individual students to facilitate student success. Assessment may include, but is not limited to, information regarding the students' study skills, English language proficiency, computational skills, aptitudes, goals, learning skills, career aspirations, academic performance, and need for special services. Assessment involves the collection of such information at any time before or after enrollment, except that the process of assigning a grade by an instructor shall not be considered part of the assessment process. Once a grade has been assigned and recorded in a student's transcript it can be used in the assessment process. (Section 55502[b])**

The intent of the definition is to be inclusive rather than exclusive in terms of what constitutes assessment in matriculation and therefore what must undergo review. The review focus is on the mechanism for gathering information, i.e., the instruments, methods and procedures that are employed. The mechanisms of assessment include but are not limited to:

> **interviews, standardized tests, holistic scoring processes, attitude surveys, vocational or career aptitude and interest inventories, high school or college transcripts, specialized certificates or licenses, educational histories and other measures of performance. The term "assessment instruments, methods or procedures" also includes assessment procedures such as the identification of test scores which measure particular skill levels, the administrative process by which students are referred for assessment, the manner in which assessment sessions are conducted, the manner in which assessment results are made available, and the length of time required before such results are available. (Section 55502 [c])**

Furthermore, the regulations provide guidance in that specific practices are prohibited. In implementing matriculation services, community college districts shall not do any of the following:

> **(a) use an assessment instrument which has not been approved by the Chancellor pursuant to Section 55524, except that the Chancellor may permit limited field-testing, under specified conditions, of new or alternative assessment instruments, where such instruments are not used for placement and are evaluated only in order to determine whether they should be added to the list of approved instruments (Section 55521[a]);**
>
> **(b) use any assessment instrument in a manner or for a purpose other than that for which it was developed or has been otherwise validated (Section 55521[b]);**
>
> **(c) use any single assessment instrument, method or procedure, by itself, for placement, required referral to appropriate services, or subsequent evaluation of any student; provided however that, in the case of assessment instruments, the use of two or more highly correlated instruments does not satisfy the requirement for use of multiple measures (Section 55521[c]);**
>
> **(d) use any assessment instrument, method or procedure to exclude any person from admission to a community college (Section 55521[d]);**
>
> **(e) use any assessment instrument, method or procedure for mandatory placement of a student in or exclusion from any particular course or educational program, except that districts may establish appropriate prerequisites pursuant to Sections 55002 and 58106 (Section 55521[e]); or**
>
> **(f) use any matriculation practice which has the purpose or effect of subjecting any person to unlawful discrimination prohibited by Chapter 5 (commencing with Section 59300) of Division 10 of this Part. (Section 55521[f]).**

These regulations provide the context for establishing the procedures and standards for review. The implication is that any information gathered about an individual student which is subsequently used in the matriculation process by the student or by others to make decisions about the student other than for

the assignment of a grade falls under the definition of assessment and thus, must be reviewed.

Very importantly, some prohibited assessment practices involve consequences resulting from the use of a test. Evidence that these negative consequences do not result from the use of the assessment mechanism in question must be provided. As a specific example, the Chancellor is charged with ensuring "... that all assessment instruments included on the list minimize or eliminate cultural or linguistic bias..." Disproportionate impact resulting from the use of assessment is the issue being addressed. From the regulations, disproportionate impact is defined to occur when:

> **the percentage of persons from a particular racial, ethnic, gender, age or disability group who are directed to a particular service or placement based on an assessment instrument, method or procedure is significantly different than the representation of that group in the population of persons being assessed and that discrepancy is not justified by empirical evidence demonstrating that the assessment instrument, method or procedure is a valid and reliable predictor of performance in the relevant educational setting.**

The regulations require documentation and evidence addressing this issue:

> **(a) Each community college district shall establish a procedure for ongoing evaluation of its matriculation process to ensure compliance with the requirements of this chapter. (Section 55521[a])**
>
> **(b) As part of the evaluation required under subsection (a), all assessment instruments, methods or procedures shall be evaluated to ensure that they minimize or eliminate cultural or linguistic bias and are being used in a valid manner. Based on this evaluation, districts shall determine whether any assessment instrument, method or procedure has a disproportionate impact on particular groups of students described in terms of ethnicity, gender, age or disability, as defined by the Chancellor. When there is a disproportionate impact on any such group of students, the district shall, in consultation with the Chancellor, develop and implement a plan setting forth the steps the district will take to correct the disproportionate impact. Community college districts shall also evaluate the impact of assessment policies on particular courses or programs. (Section 55521[b])**

Evidence on disproportionate impact, therefore, must become part of the criteria used at some level in reviewing assessment instruments, methods and procedures. Information used to make decisions that affect individual students, the instruments, methods and procedures for gathering and the consequential use of the information must be reviewed except for individual course assessments used as part of the course instructional design or for the assignment of course grades.

Similarly, another regulation that needs to be incorporated into the review and subsequent judgments on the adequacy of an assessment process is the stipulation that multiple pieces of information must be used for placement, required referral to appropriate services or subsequent evaluation of any student. The required use of more than a single assessment instrument,

method or procedure for making these decisions is found in Section 55521 [c], under the list of prohibited practices identified previously. This regulation again identifies a situation in which an instrument might be reviewed favorably at one level of review, but because it is not to be used as the only source of input for placement decisions at the local level, the total assessment process (of which the information from a single instrument is but one part) will need to be reviewed and approved or disapproved.

The Assessment Review Focus

To focus this initial review, the use of assessment for matriculation in California community colleges as a placement tool has been selected. Selecting for initial review those instruments, methods and procedures that serve the placement function is consistent with the spirit of AB3, the emerging regulations, and sound educational practice. Assessment resulting in appropriate, i.e. valid, placement of students in courses and programs will serve the desired outcome of matriculation, which is to "facilitate student success in college." Use of assessment for placement purposes is also the dominant practice in California community colleges as indicated by surveys of assessment practices.

To select instruments in the current review, two criteria were used to define the placement function in matriculation. Any assessment instrument, battery, or device used in one of the following manners was intended for inclusion:

a. The instrument battery or device is used to assist/help with the appropriate placement of students in different levels of instruction (e.g., reading, writing, mathematics), classes or programs.

b. The instrument, battery, or device is used in advising students on course selection, career choice/path, or personal guidance/counseling.

In reviewing instruments that assist with the placement function, the distinction again needs to be noted between the potential for an instrument to validly serve its function and meet the mandate of AB3 versus the actual use of information from the instrument and the consequences of its use at the local level. A requirement of AB3 is that accurate (reliable and valid) information be used in forming the placement recommendations.

SECTION II: STANDARDS FOR THE EVALUATION OF ASSESSMENT MEASURES

The Identification of Standards

The Standards for Educational and Psychological Testing (1985) is a document specifying guidelines for the development and use of tests. The **Standards**' purpose is "to provide criteria for the evaluation of tests, testing practices, and

the effects of test use" (Standards, p. 2). Prepared jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, the document represents a considerable effort on the part of many psychometric experts and test users to develop criteria for the development and evaluation of assessment instruments. The **Standards** is intended to provide a basis for evaluating the quality of testing practices as they affect the parties involved. Accordingly, the standards will serve as the primary reference for the evaluation of measures in use by the California community colleges.

Since the philosophy underlying this document is similar to the one underlying the **Standards**, reviewers of tests for community colleges must become familiar with the information in the **Standards**. In general, the **Standards** state that the evaluation of tests should ultimately involve judgment and not some mechanical process such as a checklist. This judgment is to be based on a general knowledge of the behavioral sciences, a specific knowledge of the professional field to which the test applies, and alternative measures that are available for the same purpose. Also, knowledge of psychometrics along with a keen sense of practical issues such as feasibility of test use should play a role in the evaluation of any measure.

Introduction to the Standards for the Evaluation of Measures in Use by the California Community Colleges

The purpose of the **Standards** is broader than the purpose of this document: the establishment of criteria for the evaluation of assessment measures for the California community colleges. Accordingly, the following is an abbreviated version of the **Standards**. This abbreviated version reorganizes the **Standards** into five sections:

1. Validity;
2. Reliability and Errors of Measurement;
3. Scaling, Norming, Score Comparability and Equating;
4. Standards for Administration, Scoring and Interpretation; and,
5. Testing Special Groups.

Each section contains a number of criteria, with each criterion grouping together a number of similar principles from the **Standards** that have been paraphrased. Each section identifies those standards that have been culled for that presentation. This abbreviated version of the **Standards** is presented for the convenience of the test reviewers for the California community colleges. The standards that have been abstracted are those that are expected to relate frequently to the community college reviewer's needs. However, test reviewers should be familiar with the **Standards** in its entirety and rely on its content as the primary source.

A related document that affords considerable guidance in the review of assessment devices is the Code of Fair Testing Practices in Education (1988) prepared by the Joint Committee on Testing Practices and distributed by the National Council on Measurement in Education. Unlike the **Standards**, which

is designed to address the full range of testing applications, the Code is intended to be consistent with the **Standards**, yet differs in both audience and purpose. While based on the **Standards**, the Code has been prepared for the public at large. The Code narrows its scope to educational testing and focuses on those criteria that effect the proper use of tests. Given the Code's specific attention to educational testing, this document has been relied on as well to guide the identification of those criteria.

A third primary source for detailing test standards judged particularly relevant to California community colleges was the Equal Employment Opportunity Commission's (EEOC) Uniform Guidelines for Employee Selection Procedures (1978). While prepared to provide guidance in matters of the use of tests for the selection of employees, the EEOC Guidelines do identify numerous criteria that can lead to the fair and equitable use of testing. As such, when specific criteria were evaluated for inclusion in the abbreviated version of the Standards, compliance with practices suggested in the Guidelines was monitored.

The criteria that are presented are compatible with these sources, but were identified specifically for evaluating assessment measures used by California community colleges. These criteria are offered for a number of related reasons. The **Standards** are intended to offer guidelines for test developers or test users who may be working with one of a number of types of tests used in one of a variety of settings. Therefore, the **Standards** are very broad and allow for multiple exceptions. In contrast, the criteria for the community colleges can be more specific. For example, since the California community colleges consist of students from diverse populations, the criteria can identify as a requirement that test users must document how they plan to meet the special needs of these populations. Second, by having more specific criteria, the test reviewers should make more reliable judgments. Accordingly, the review system is more likely to be viewed as fair in that a test should be evaluated from the same basis regardless of the judges.

Finally, evidence for some criteria may likely not be available for review by the external test reviewers in their evaluations or are not restricted as required documentation by the test developer. However, local college personnel can expect that such evidence will be evaluated by the Chancellor's office during on-site visits to determine that such criteria are met. These criteria are specific to a particular college, such as criteria dealing with cut scores.

Select Criteria for the Evaluation of Measures in Use by the California Community Colleges

Section 1: Validity

Validity is the most important area for evaluating a test. Evidence of validity speaks to the suitability of the specific interpretation attached to test scores. Seen in this light, the test itself is not validated. Rather, it is the use made from the test information that is to be validated. If use or practice differs across applications, then each specific application needs validation. Validity

criteria that are especially germane to the California community colleges' use of tests follow. Unless otherwise noted, the test developer ordinarily has responsibility to provide the information called for by the criterion. Specific references to items in the **Standards** are given for each criterion.

Criteria

Criterion 1. Validity: General. (Standards 1.1 to 1.4, 3.1 and 6.1 to 6.3) Evidence should be provided by the test developer, test researchers, or users supporting the particular use(s) of the test. Choice of types of evidence provided should be specified as well as a rationale for the mix of evidence. The specified purpose determines whether the reviewer should selectively weigh construct-, content-, or criterion-related evidence in the evaluative process.

Situations may occur in which, in a user's judgment, decisions may be based in part on tests for which little evidence of validity for the intended purpose is available. In these circumstances, the user should take great care not to imply that the test has established validity or place considerable reliance on the measure in decision making.

Necessary cautions concerning lack of evidence for the use of a measure, subscore, item, difference score, or profile interpretation are required within the testing materials.

If a user alters a test in some fashion, the user needs to revalidate the test under the changed condition or offer a rationale for why revalidation is unnecessary. Any substantial change in a test's format, alteration of how the test is administered, its language, instructions or content requires that additional validation evidence be assembled. The extent of the additional evidence required will depend on the nature of the change (s).

Criterion 2. Content-related Evidence. (Standards 1.6, 1.7, 3.3 and 10.6) A clear definition of the universe represented, its relevance to the test's use, and the procedures used to relate the item content to the universe should be described fully and accurately. Sufficient detail is expected such that users can evaluate the range of content in the measurement as appropriate for a domain. Even if the content-related evidence is fully described, a negative judgement might be rendered if the information does not support the choice of test content.

If expert judges are used to make content evaluations, their qualifications must be specified.

Criterion 3. Construct-related Evidence. (Standards 1.8 to 1.10) If a test is proposed to measure a construct, the construct should be specified and well defined. In addition, evidence to support the inference from the test to the construct should be presented. Evidence is appropriate to show not only that the instrument correlates with conceptually related measures, but also that it does not correlate with conceptually unrelated measures.

Criterion 4. Criterion-related Evidence: General. (Standards 1.11 to 1.18, and 1.25) All criterion-related studies should be completely described, including a specification of the sample, statistical analyses, criterion measures, and the time intended between predictor and criterion measurements. In addition, any factors that affect the results of the study should be reported. The criterion-related evidence should be judged to determine its support for and relevance to the intended use of the test. Further, criterion measures need to be described and the rationale for their selection provided. The generalization of the validity study to the intended use must be described.

Criterion 5. Criterion-related Evidence: Differential Prediction. (Standards 1.20 to 1.22, and 8.10) Differential prediction should be investigated under two conditions: (1) when feasible, and (2) when prior research has established a substantial likelihood for differential prediction to occur with a particular type of test. That is, when groups differ in their demographics, past experiences, or instructional treatment and such factors are related to performance on the test, then investigations must be undertaken to determine if decisions are systematically different for the members of a given group than for all groups combined.

Any differential prediction studies should involve proper statistical and methodological considerations.

Criterion 6. Cut Scores. (Standards 1.23 and 6.9) In some applications, users reach one decision if a test taker scores at or below one score (the cut score) and a different decision if the test taker scores above that value. The test user is required to document the method, rationale, and appropriateness for setting that cut score. Evidence of the appropriateness of the cut score for the decision must be documented.

When cut scores are based on professional judgment, the qualifications of the judges should be documented.

Section 2: Reliability and Errors of Measurement

Error in measurement is inevitable. Recognition of this fallibility serves as its own caution and requires that the likely extent of error associated with test scores be documented. While many (potential) sources of error exist, an essential requirement is that the approach to documenting the extent of measurement reliability takes into account errors of greatest concern for a particular test use and interpretation. Criteria associated with the reliability of tests most germane to California community college uses are presented below.

Criteria

Criterion 1. Estimates of Reliability and Standard Errors: General. (Standards 2.1 to 2.5, and 2.9) For each reported score, estimates of reliability and standard errors of measurement should be provided to determine whether

scores are sufficiently accurate for the intended use. The sample characteristics, statistics, and, more generally, the methodology employed to document a test score's reliability should be described completely. If theoretical or empirical reasons exist to suggest that estimates differ by population, estimates should be provided for each major population.

Even if appropriate estimates are provided and the employed methodology is acceptable, the measure should be evaluated negatively if the estimates of reliability and standard error indicate low accuracy for the test's intended use.

Situations may occur in which, in a user's professional judgment, decisions should be based in part on tests for which little evidence of reliability is available. In these circumstances, the user should take great care not to imply that the test has established reliability.

Criterion 2. Estimates of Reliability and Standard Errors: Type of Estimates Provided. (Standards 2.6 to 2.8, and 3.13) The type of reliability estimate should be carefully selected, appropriate for the expected use, and properly interpreted. For example, coefficients of internal consistency should not be interpreted as estimates of stability over time. Coefficients that yield spuriously high estimates of reliability for speeded tests should not be used, if total performance is dependent on test takers' inability to complete a test due to time constraints. When corrected coefficients are reported, uncorrected indices must as well be presented. If a test is scored using a judgmental process, the degree of error introduced by scoring should be documented.

Criterion 3. Estimates of Reliability and Standard Errors: Specific Applications. (Standards 2.10 to 2.12, and 11.3) For tests whose use relies on cut scores, reliabilities need to be reported at the cut score or for score intervals. Also, decision consistency reliability information needs to be reported for select score points. For adaptive tests, reliabilities must be reported using repeated administrations using different items.

Section 3: Scaling, Norming, Score Comparability and Equating

The metric used to report test scores is chosen to enhance the interpretability of the information shared with the user or test taker. In the same vein, access to norms provides a means of referencing performance to a defined population of persons or groups. The utility of having multiple forms of a test available is realized only when performance can be reported on a single common scale. (Otherwise there will be an advantage or disadvantage associated with the particular form taken). Regardless of the utility achieved by establishing derived scales, resultant transformed scores can introduce error to the measurement process due to the procedure itself or the sampling methodology used. Important criteria for evaluating the appropriateness of test score transformations used in the community colleges are presented below. As these criteria suggest, benefits realized by transforming scores must be carefully weighed.

Criteria

Criterion 1. Choice of Scales. (Standards 4.1 and 4.2) The method used to compute the transformed (derived) scale or raw score should be clearly delineated. In addition, the rationale should address the relationship between the scaling methodology and the test's purpose. The measure is strengthened by the transformation of scores to the degree that the choice of scores is appropriate for and consistent with the intended purpose.

Criterion 2. Norms. The group of persons used for establishing the norm should be well described and appropriate for the test's intended use. The norm group must be groups to whom users will wish to compare the individuals tested. The methodology for the constructing norms, including sampling plan, participation rates, and descriptive statistics, should be exhaustively specified. In addition, the year(s) that the data were collected should be reported. Out-of-date norms should be avoided.

Criterion 3. Comparability of Scores. (Standards 4.6, 4.7, 4.9, and 5.9) When scores based on different test forms or different response formats are intended to be interchangeable, data that supports the equivalence of the forms must be documented. If the content of the instrument changes across years, but the scale is intended to be comparable, the method for maintaining comparability should be described and be adequate for its intended use. When the test specifications change across years, the change should be fully described, the rationale for the change given, and the test user informed of the extent to which scores remain interchangeable.

Section 4: Standards for Administration, Scoring and Interpretation

Information that documents the properties of a test must be available to users. In this respect, manuals and technical reports are particularly important for detailing how to appropriately use a particular measurement device. Completeness, accuracy, and clarity remain key to the proper use or opportunity to select a test. Criteria that speak to the adequacy of communication and use the of information for community college staff are presented in this section.

Criteria

Criterion 1. Administration and Scoring. (Standards 3.21, 3.24, 5.4, 5.10, 6.6, 8.5, 15.1 and 15.4) The standardized procedure(s) for the administration and scoring of a measure should be fully described in the test's manual. A test manual should identify the qualifications necessary to administer the test appropriately. The standards for test administration and scoring specified in the manual should match the characteristics of the test. Any modification of standard test administration procedures or scoring should be fully described in the manual with appropriate cautions noted. An agency that develops a test has the same obligation to supply manuals and technical reports as does a commercial test publisher.

Criterion 2. Interpretation: General. (Standards 5.4, 6.12, 6.13 and 8.3) A test manual should identify necessary qualifications to interpret test results. Tests should not be interpreted as ability tests without considering alternative explanations. Screening measures should be used only for identifying individuals for further evaluation. Test users should not use interpretations of test results unless they have documentation that indicates the validity of the interpretations for intended use and on the samples on which they were based.

A test user that makes educational decisions based on differences in scores, such as aptitude and achievement scores, should take into account the overlap between the constructs and the reliability or standard error of the difference score.

Criterion 3. Interpretation: Test for Certification. (Standards 8.4, 8.7 and 8.8) If a test is used to certify completion of a given education level or grade level, both the test domain and the instructional domain at the given grade or education level should be described in sufficient detail so that the agreement between the test domain and test content can be assessed. Also, the test should not cover materials that a student has not had an opportunity to learn. Students should have multiple opportunities to take such a measure.

Criterion 4. Test Materials. (Standards 3.22, 5.7, 8.11 and 15.12) The testing materials should be readable and understandable. Materials should limit claims of test properties and characteristics to those conditions for which data exist to support the claim. Such support data should be reported.

Criterion 5. Decision making. (Standards 6.4, 6.12 and 8.12) A decision or characterization that will have a major impact on a test taker must not be made solely or automatically on the basis of a single score. Decisions are to be made in conjunction with other test information, previous classroom performance, and opinions by advisors from discussions with students.

Section 5: Testing Special Groups

Recognition of the difficulties associated with using tests appropriately with groups with linguistic differences merit careful evaluation and study if assessment results are to be valid. Also, allowing provisions for testing persons who have handicapping conditions will improve the appropriateness of the measurement. Criteria that enhance measurement and subsequent evaluation for diverse populations are offered below.

Criteria

Criterion 1. Testing Special Groups: General. (Standard 6.10) Test administrators and users should not attempt to evaluate test takers whose special characteristics, ages, disabilities, or linguistic, generational, or cultural backgrounds, are outside their range of experience. A test user should seek consultation regarding test selection, necessary modifications of testing procedures, and score interpretation.

Criterion 2. Test Design for Non-native English Speakers. (Standards 13.1 to 13.4, and 13.6) Tests, their items and their administration instructions should be designed to minimize threats to validity and reliability that may arise from language differences. Any recommended linguistic modification should be described in detail. When a test is to be used with linguistically diverse test takers, information should be provided for appropriate use and interpretation. Translated tests should be evaluated for reliability, validity, and comparability with the English-version.

Criterion 3. English Language Proficiency Tests for Non-native English Speakers. (Standard 13.7) English language proficiency should not be determined solely with a test that demands only a single linguistic skill. The caution here is not limited to the test that is dependent on a single format (for example, a multiple-choice, paper and pencil device). The attention of this standard is focused on the breadth and depth of the construct being appraised. Users need to be aware of the needs and demand a complete range of language skills, that is for written as well as reading, oral and listening proficiency.

Criterion 4. Test Design for People Who Have Disabilities. (Standards 14.1 to 14.6) Expertise, both psychometric and training or experience with populations with disabilities, is a prerequisite to modification of a test for an individual or group with disabilities. Knowledge of the effects of various handicapping conditions on test performance is essential. Until validity data are obtained for scores secured from non-standardized testing conditions, documentation must be available and the results must be interpreted cautiously. Pilot testing of the modified measure(s) is strongly advised with persons having the same or similar disability. When feasible, time limits should be modified for the person with disabilities based on reliability and validity studies conducted.

Criterion 5. User responsibility. (Standards 13.1, 14.7, and 14.8) Knowledge of alternative measures is a precondition to test selection used with linguistic minorities or persons who have disabilities. When modified forms are available, they are to be used with these persons. Proper selection of appropriate norms to facilitate score interpretation is essential. Using personnel for test administration who have been specifically trained for the group or person to be tested is strongly encouraged.

SECTION III: SPECIFIC CRITERIA FOR TEST USAGE FOR THE CALIFORNIA COMMUNITY COLLEGES

The preceding section summarizes and abstracts the portions of the Standards for Educational and Psychological Testing (APA, 1985) that are most relevant to the evaluation of measures used by the California community colleges. The **Standards** were written to help individuals assess the quality of all types of instruments and to describe the obligations of all parties involved in the testing process, including test developers and users. Because it offers standards for diverse measures and all interested parties, the **Standards** are written very

broadly and often lack specificity for a particular application. For example, although the **Standards** clearly emphasize that validity evidence is required for the use of a test, the **Standards** do not suggest what type of validity evidence is needed for a test with some particular use and the required strength of the evidence.

Most tests used by California community colleges are intended to help students select the appropriate courses to take. The tests serve a placement purpose, offering students guidance as to whether they should enroll in a course at the beginning of the sequence of courses in mathematics, for example, or somewhat later in the sequence. Because of the specific nature of these measures and their common use within the California community colleges, more explicit criteria can be written for both the test developer and the test user than what are or could be presented in the **Standards**.

More explicit criteria for the test user and test producer, referred to as the specific standards for the California community colleges or simply as the specific criteria, are needed and are desirable for a number of reasons. First, the diverse audiences involved in the California community colleges should understand the evaluation system used by the psychometric experts. In this sense, the **Standards** by themselves are too vague for establishing an evaluation policy. More specific criteria will allow all parties involved to understand what is required for the Chancellor to approve the use of a measure. Second, with more specific criteria, test developers and users will be better able to estimate whether their tests need further refinements and documentation before being submitted for review. Accordingly, fewer tests will be reviewed and the tests that are reviewed are more likely to receive a favorable evaluation. Third, the decisions that are reached by the psychometric experts should be more consistent from year to year. Without specific criteria, the likelihood is higher that the same instruments judged acceptable one year might be judged unacceptable another year because the interpretation of the Standards would vary across years. Fourth, the evaluation system is likely to be less arbitrary with explicit criteria. Finally, defining specific test criteria at the state level indicates a clear commitment on the part of the Chancellor's Office that tests be used to aid students in making important academic decisions rather than preventing them from reaching their educational goals.

The expectation is that all specific criteria will be met. However, a test developer or user occasionally might not fulfill one or more of the specific criteria, but can still justify the use of the test. A test developer or user may indicate in writing why a specific criterion has not been met. Whether an exception is granted will be considered on a case-by-case basis.

Specific criteria are being developed for four types of instruments: multiple-choice and other objective achievement measures developed by an external agency (not associated with the California community colleges); multiple-choice and other objective achievement measures developed by California community college (s); writing samples and other measures with open-ended responses; and non-achievement instruments used for placement, such as

interest measures. These specific criteria are being developed as technical appendices to the *Standards, Policies and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges.*

The minimum standards for the test developer and the test user are described below. Criteria that are the responsibility of each of these groups are presented separately. An attempt was made to describe each criterion succinctly, without attending to specific conceptual or methodological problems that might arise in the evaluation of a particular test. An assumption is that individuals have conducted all aspects of the study or analysis properly. Finally, the appropriate use of a test needs to be considered in light of all applicable standards described in the Standards for Educational and Psychological Tests (APA, 1985). Consequently, meeting these minimum standards is a necessary, but not sufficient, condition to receiving a favorable recommendation for a test's use.

SECTION IV: THE PROCESS FOR REVIEWING ASSESSMENT MEASURES IN THE CALIFORNIA COMMUNITY COLLEGE SYSTEM

Two broad categories of instruments appear to be in use in California community colleges to serve the placement function. One category contains what are typically labeled as cognitive domain instruments - assessment instruments intended to appraise the extent of mastery of learned knowledges and skills. The other category consists of those non-cognitive inventories such as interest and personality-type indicators. All the reviews prior to July, 1991 will be restricted to the placement cognitive domain instruments.

The review of instruments to be used on California community college campuses is a responsibility shared among the Chancellor's Office, the colleges, test developers, and users of the particular test as well as agents or agencies contracted to provide specific review and evaluation services. In this section the steps are presented that identify the process by which assessment instruments are reviewed. The responsibilities of the various parties are discussed.

Understandably in the final analysis, the Chancellor of the California community colleges is vested with the decision to allow or disallow use of any instrument on the campuses. Procedurally, the Chancellor, in coming to a specific decision, seeks the advice of the Matriculation Advisory Committee (MAC) and this group relies on its Assessment Work Group to guide MAC's recommendations to the Chancellor. In addition, the Chancellor's Office will assume responsibility to communicate instrument evaluation results by the psychometric experts to the appropriate testing agency responsible for the production and distribution of the instrument.

The evaluation results are to be based on those professional standards that guide educational and psychological testing, and those conditions identified in AB3 and the matriculation regulations. While the instrument review steps are discussed in terms of activities leading to a recommendation to the Chancellor, the process can also be adopted by a local district for arriving at its

own decisions regarding a particular assessment instrument or process. The approach and procedures for the review of instruments are as follows.

Step 1. Compile Information on Assessment Instruments. The quality of the recommendation made by the MAC to the Chancellor depends upon the quality of the information available to it. Consequently, as much information as possible should be available to the committee, including but not limited to, the test and its manuals, technical reports, reviews such as those found in the Mental Measurement Yearbooks, articles that review a test published in professional journals and books, and technical reports prepared by California community college users. Assembling this information is the responsibility of the psychometric experts. Test developers are to be contacted and asked to provide copies of tests, test documents and related technical reports.

Although much of the documentation on a measure is likely to be supplied by the instrument's developer and made available in a manual, the user has the responsibility for supplying information that indicates that a measure is being used appropriately. Thus, the California community colleges are to be contacted and asked to provide copies of technical reports they have prepared.

Step 2. Perform Psychometric Expert Review. The information gathered in step 1 will be reviewed by at least two psychometric experts. The psychometric experts must have received their doctorates in a measurement-related area or have had five or more years experience in an occupation requiring expertise in tests and measurement. They must have a broad understanding of both theoretical and applied issues associated with testing so that they may make informed recommendations to the Chancellor's Office. These experts will evaluate a measure based primarily on criteria presented in the Standards for Educational and Psychological Testing, as well as the evaluation criteria that have been prepared specifically for the evaluation of instruments for the California community colleges. Reviewers may also use other guidelines that are commonly accepted by the psychometric community such as the Code of Fair Testing Practices in Education. In addition, the measure must be reviewed to ensure its compliance with the matriculation regulations. Selection and defining the workscope of the psychometric experts is a responsibility of the Chancellor's Office.

Step 3. Perform Content Expert Review. The test documentation gathered in step 1 also will be reviewed by at least two subject matter content experts. For example, if a measure purports to assess "Preparedness for Calculus," individuals who understand the prerequisite information for learning calculus and the information that is presented in a calculus course would be solicited to review the information concerning the test as well as the test itself. These reviewers will be sampled from lists of potential content experts recommended by members of the MAC Assessment Work Group as being knowledgeable about specific content area courses for which an assessment instrument is to have value. To the extent possible, content expert reviewers will be chosen to reflect the diversity of experiences with a given instrument in the colleges. The content reviewers' primary focus is to evaluate: 1) the match among the rationale underlying the measure as stated in the manual or by the

user, the items on the measure, and the suggested interpretation of scores relative to the test's its intended use, and 2) the content appropriateness for the diverse populations served in the California community colleges. The subject content experts are necessary because in most instances the psychometric experts will have limited knowledge of the necessary and essential information to make a decision about test content, whereas the content experts should be able to offer a more informed opinion.

The content experts will file a written report with the psychometric experts so that they may have as complete an understanding of the quality of the measure as possible. A review form will be utilized to guide and standardize the content expert's review.

Step 4. Perform Matriculation Advisory Committee (MAC) Assessment Work Group Review. The MAC Assessment Work Group will consist of individuals who work within the California community colleges. Their charge is to serve in an advisory capacity to the psychometric experts, the Matriculation Advisory Committee and the Chancellor's staff. Initially this Group will assist and be consulted in establishing the formal standards, policies and procedures for the review of tests used in the California community colleges. The members should include a cross-section of individuals who have expertise in assessment, research and evaluation or testing, are responsible for administering, scoring tests and interpreting tests, are faculty members of community colleges, or are community college administrators. Given their membership, they can also serve as a liaison with the field. Once the procedures for the instrument review have been finalized, the Work Group is to be consulted and informed regarding the actual evaluation of measurements. The psychometric experts are to present their findings and judgments to the Work Group. Members of the Work Group will be asked to offer their opinions of the evaluations. Since the members have an understanding of how assessment measures are used within the community colleges they should be able to give feedback about the evaluation of an assessment measure as it applies to decisions made at the community colleges. As the Work Group finds it necessary, it may solicit additional information from test developers or test users.

Step 5. Generate Recommendations. For each test evaluated, a written report is to be filed by the psychometric experts with the Chancellor's Office. Prior to filing a recommendation with the Chancellor regarding a specific test, an Interim Test Evaluation will be shared with the test publisher. The interim filing with the test producer will give the developer the opportunity to be informed about the report, and, if elected, to respond to it. A period of fourteen days following the interim filing will be allowed for the test producer to provide additional information that might lead to modification of the recommendation report. The interim period is not planned for the developer to assemble, analyze and report on "new" data gathered in response to the preliminary findings; rather, the interval provides an opportunity for the test developer to supply information that is already available, but not previously provided for the review.

The report filed by the psychometric experts should include a brief summary of the opinions of the content experts, the MAC Assessment Work Group, and available information gleaned from external test reviews, test publishers, and the psychometric experts for the measure. In addition, the report should offer a recommendation concerning the use of the measure. A recommendation will be in one of four categories: (A1) Approval; (A2) Provisional Approval; (A3) Probationary Approval; or (B) Disapproval. The first three categories are intended to communicate different levels of approval with different consequences attached to the recommendation. Any instrument placed in one of the three "A" categories will be available for use by a community college. The length of time the instrument will be available without submission of additional information varies by category. The intended implications for the use of an instrument when placed in a category are as follows.

A1 ***Approval*** - Instruments in this category meet the standards' criteria. The available evidence indicates these instruments have potential value when used to serve a specific assessment function in California community colleges. These instruments have high probability of yielding test scores useful in assisting decision making for a particular community college student.

A2 ***Provisional Approval*** - Instruments in this category meet relevant standards and criteria, but lack sufficient or recent information to assign the unequivocal "Approval" rating. Some criteria were not met because (a) documentation was lacking that in all likelihood should be provided in a relatively short time period or (b) criteria were recently introduced and the user/developer should be allowed time to meet them. In this conditional category the expectancy is that the test, in time, will achieve an "Approval" recommendation. As such, the necessary clarifying information to receive an "Approved" evaluation on instruments for this "Provisional" category is expected to be provided in due course. Recommendation in this category means that the test developer or user must supply within one academic year the specified additional clarifying information. Failure to submit the required clarifications within one year will result in reclassification into the "Probationary" category.

A3 ***Probationary Approval*** - Instruments in this category are missing critical information and thus a clear-cut recommendation cannot be established; or from the information that is available, deficiencies are noted. The intended purpose for use of these instruments is clearly stated and some positive information supporting its use is available, but the necessary evidence available for a final judgment is incomplete. For tests that are recommended as "Probationary," additional data collection must be provided for further evaluation. Such instruments can only be placed into an "Probationary Approval" designation for a maximum of <u>two years</u>.

B *Disapproval* - Instruments in this category are those where the evidence indicates that they have failed to meet one or more of the standards or criteria considered essential by the reviewing bodies or fail to meet a condition of AB3. What is considered an essential element is likely to vary among applications (that is, tests can be disapproved for differing reasons), but the specific deficiency will be identified in the report of the experts to the Chancellor.

Step 6. Disseminate Chancellor's Decision. The Chancellor will make a decision concerning the use of each measure reviewed. This decision may or may not be consistent with the recommendations given by others who participated in the review process. These decisions will be communicated formally to both the community colleges and the specific test publishers involved.

Step 7. Identify Appeals Process. A recommendation by the Chancellor may be appealed, that is, subject to formal reconsideration, by any individual, college, agency or entity. Requests for an appeal are to be submitted to the Chancellor and must clearly delineate why the decision is being questioned. The Chancellor's prerogative is to determine the next course of action, although one might expect that an Appeals Committee will be called upon to reconsider the standing recommendation. Appeals are to be acted upon, and a recommendation forthcoming from the Chancellor within six months of the Appeal request.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1988). *Code of fair testing practices in education*. Washington, DC: American Psychological Association.

Equal Employment Opportunity Commission , Civil Service Commission, Department of Labor and Department of Justice. (1978). Adoption by four agencies of the Uniform Federal Guidelines on Employee Selection Procedures. *Federal Register, 43,* 38290-38315.

ERIC CLEARINGHOUSE FOR
JUNIOR COLLEGES

APR 05 1991